Blue Gene/L System Software Update

Kim Cupps

February 23, 2005

UCRL-PRES-210094

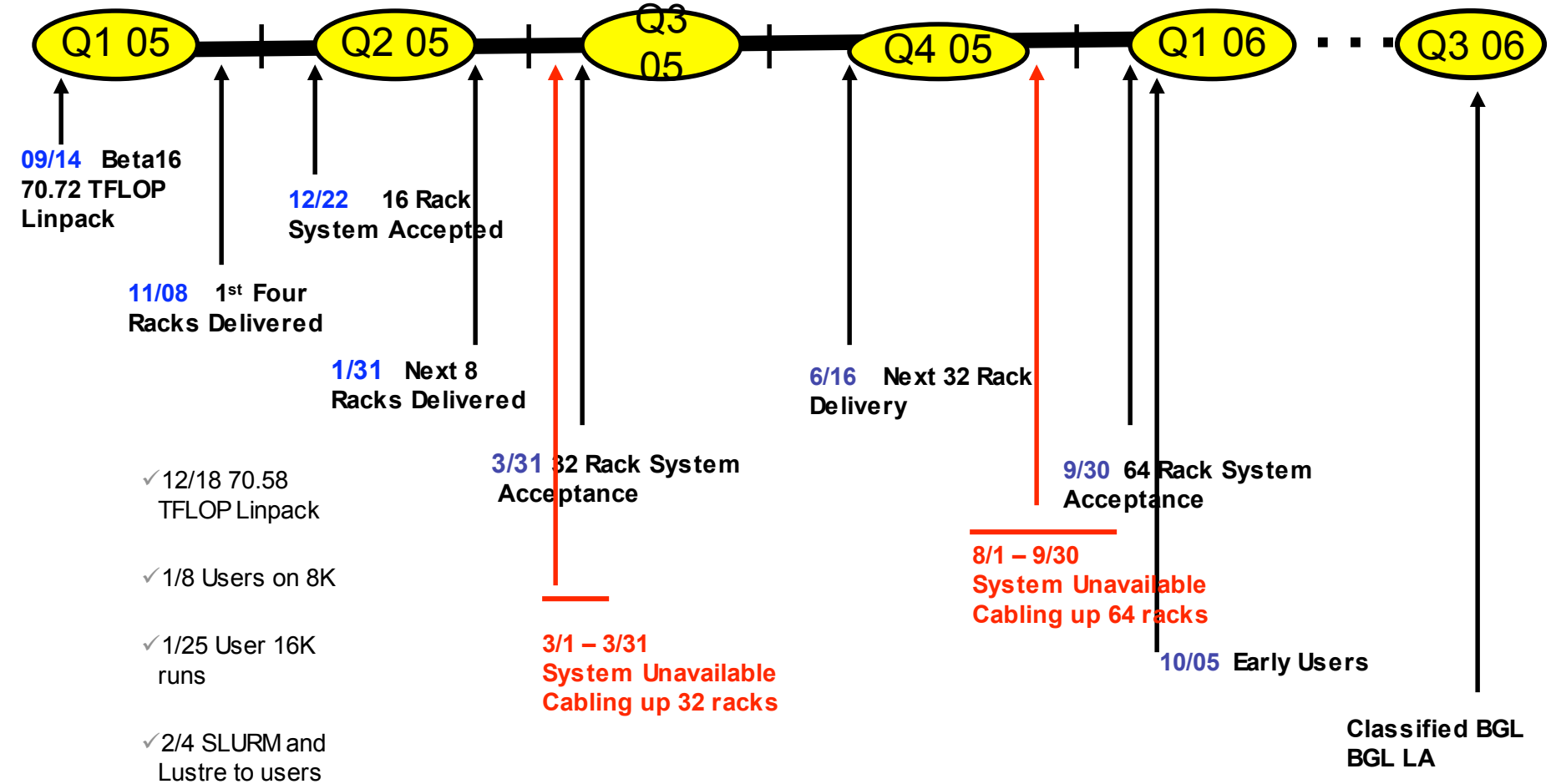
Agenda

- Current Status
- Project Timeline
- System Configuration
- Challenges
- What Lies Ahead
- System Reliability
- Early Successes

Blue Gene/L Status
Only Two More Doublings to Go

- We have 16K nodes up and running
- We have another 16K nodes on the floor cabled up, going through individual rack bringup and shakeout process
- Lustre filesystem is being used
  - NFS “pseudo parallel” filesystem will eventually disappear
- We are going to double what we have now **twice** before we finish

Blue Gene/L Integration Timeline



A Day in the Life of a BGL User

- Familiar environment
  - Login, compile and run on front end nodes
  - xlc, xlC, xlf compilers, debug using TotalView
- Collegial scheduling – everybody talks to each other via instant messaging mechanism
- We plan to run BGL with fixed size partitions that get changed twice a week
  - 16K partition is currently the maximum partition size
  - Currently we run in the (1) 16K job mode 2 days per week for scaling and software debug runs
  - 5 days a week we run 8K, 4K, 2K, and (4) 512 node partitions
- Currently all **Compute Nodes** and I/O nodes must be re-booted between jobs
  - This takes about 8 minutes for 16K partition, including mounting Lustre

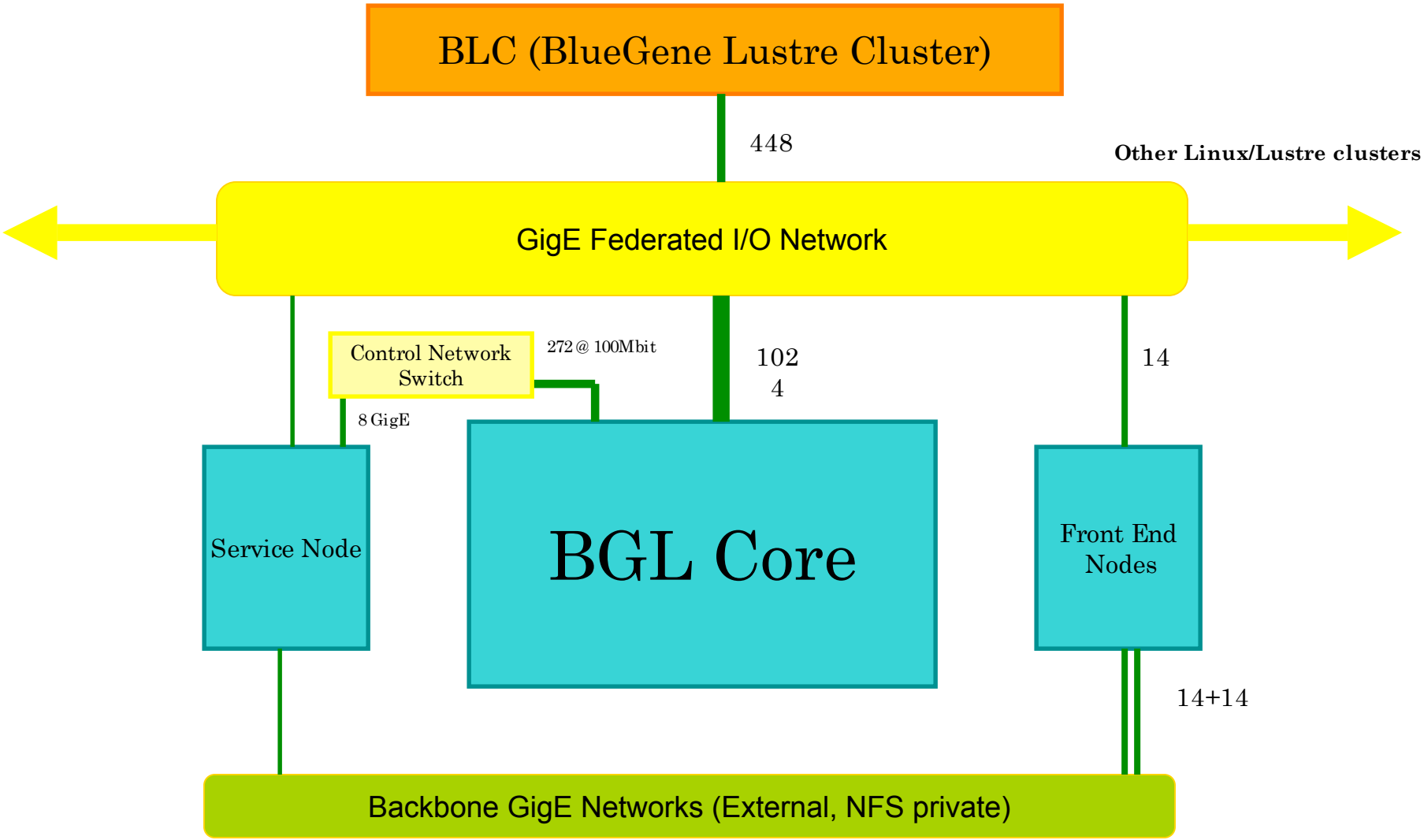
BG/L System Configuration Overview

- Blue Gene/L Core = Compute Nodes + IO Nodes
- Compute Nodes
  - 64 racks
  - 1K compute nodes per rack
- IO Nodes
  - 16 IO nodes per rack
  - 1024 IO nodes
- Compute and IO Node specs
  - dual-processor PPC 440 @ 700Mhz
  - 512 MB DDR memory
  - 4 MB L3 cache
- Compute Node : IO node ratio
  - 64:1 ratio is higher than most other BGL systems

System Configuration Overview (cont'd)

- Service node: IBM p670
  - 16 Power4+ processors, 64 GB memory
  - Runs SuSE SLES 8
- Front end nodes: BladeCenter JS20
  - Fourteen dual-processor PPC 970 blades @ 1.6GHz
  - Runs SuSE SLES 9
- BLC = Blue Gene Lustre Cluster
  - 224 dual-processor Intel EM64T "Nocona" OST nodes at @ 2.8 Ghz
  - ~900 TB Lustre filesystem
  - Target delivered I/O bandwidth of 32GB/s to user apps (32MB/s/ION)

System/Network Layout



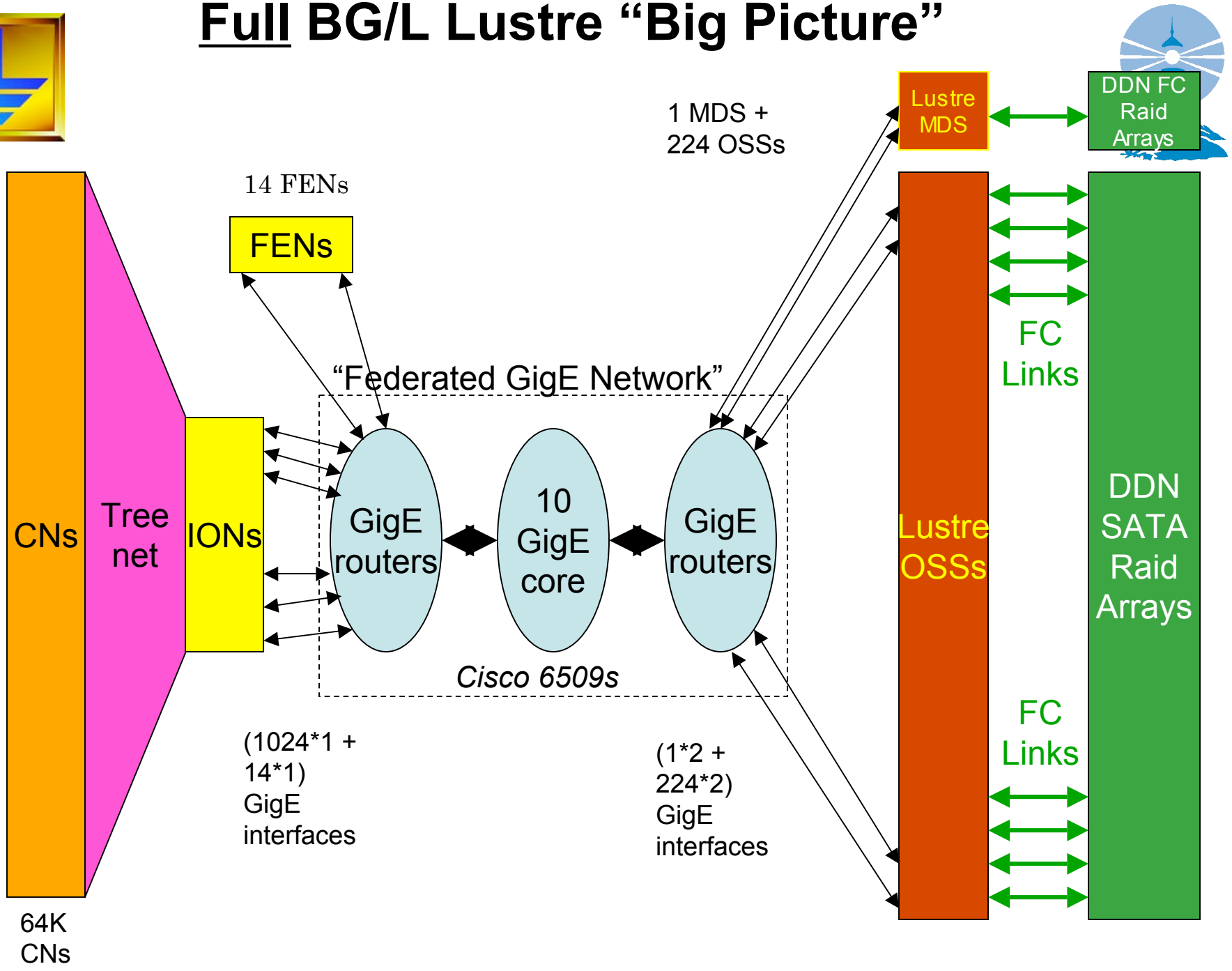
Challenges

- Port Lustre to I/O Nodes and Front End Nodes (FENs)
- Port SLURM to BG/L
- Port TotalView to BG/L
- Operational Challenges

BG/L Lustre Integration Challenges

- Server side infrastructure (224 OST nodes, 900 TB back end storage) is operational but only 1/3 of it has been used enough to shake out HW issues.
- Port of Lustre to JS20 front end nodes (PPC64, SLES9) is complete, stress testing in progress, performance looks good with SLES9 2.6 kernel.
- Port of Lustre to BGL IO nodes (PPC32, 2.4.19 kernel) is functionally complete, but many issues still need to be addressed:
  - Several issues have been debugged both in IBM and CFS software
  - Performance is improving on a weekly basis
  - Time to mount Lustre after partition reboots was slow but has been improved (< 1 minute for 16K nodes)
  - Failure of a single Lustre client on IO node requires reboot of entire partition. BGL has no provision to reboot a single IO node.
  - Scaling and stability issues are likely still ahead…

Full BG/L Lustre "Big Picture"



Resource Management Challenges

- Simple Linux Utility for Resource Management (SLURM) can define, create and destroy partitions (via *smap, slurmctld* on service node), and queue/run user jobs (via *slurmd, srun* on front end nodes)

- Outstanding issues in IBM software stack:
  - Partition must be rebooted any time a different user's job is launched.– so running static partition sizes doesn't prevent reboots.
  - Partition reboot required when changing from Virtual Node Mode to Co-Processor mode even under the same user

- LCRM ported but not tested

Operational Challenges

- Five different Operating Systems
- First experience with DB2 on Linux
- First experience with SuSE Linux
  - Need to port existing system management tools to SLES8/9
  - Lustre port to SLES9 FENs requires kernel modifications.
  - General lack of responsiveness by SuSE support
  - Figuring out a patch strategy: need to address security vulnerabilities but avoid breaking BGL software stack.
- Security issues with BGL control system software stack
  - IBM is performing an internal audit of this software.
- Support
  - IBM's BGL support infrastructure still being defined
  - Two hour IBM response time between 6 AM and 3 PM, no off-hours IBM support
  - Single point of contact via telephone
  - Problem tracking and reporting system not fully functional

Code Development and Scaling Challenges

- TotalView port is progressing well, issues are being reported to IBM
  - TV/mpirun doesn't work in virtual node mode
  - TV/mpirun doesn't work when number of tasks is less than full partition size
- MPI functionality and performance issues being tracked and reported to IBM
- Compiler issues are being tracked and reported to IBM
- Code scaling effort progressing well

System Reliability

- Hardware infant mortality rate is slightly less than one might expect
  - 6 compute node failures, 2 IO node failures since IBM install team left on 12/20.
- System Software stack is still under development
  - New drivers in development at IBM every 2 weeks
  - Major functionality changes between driver versions (including API changes)
  - Control system software is least mature

BGL Accomplishments

- 16K Linpack result achieved five weeks after delivery of the first rack.
- Functional acceptance of 16K system achieved six weeks after delivery of the first rack.
- RAS Database is easy to use and fairly robust
- Users running REAL SCIENCE on the system shortly after acceptance test completion

What's Ahead

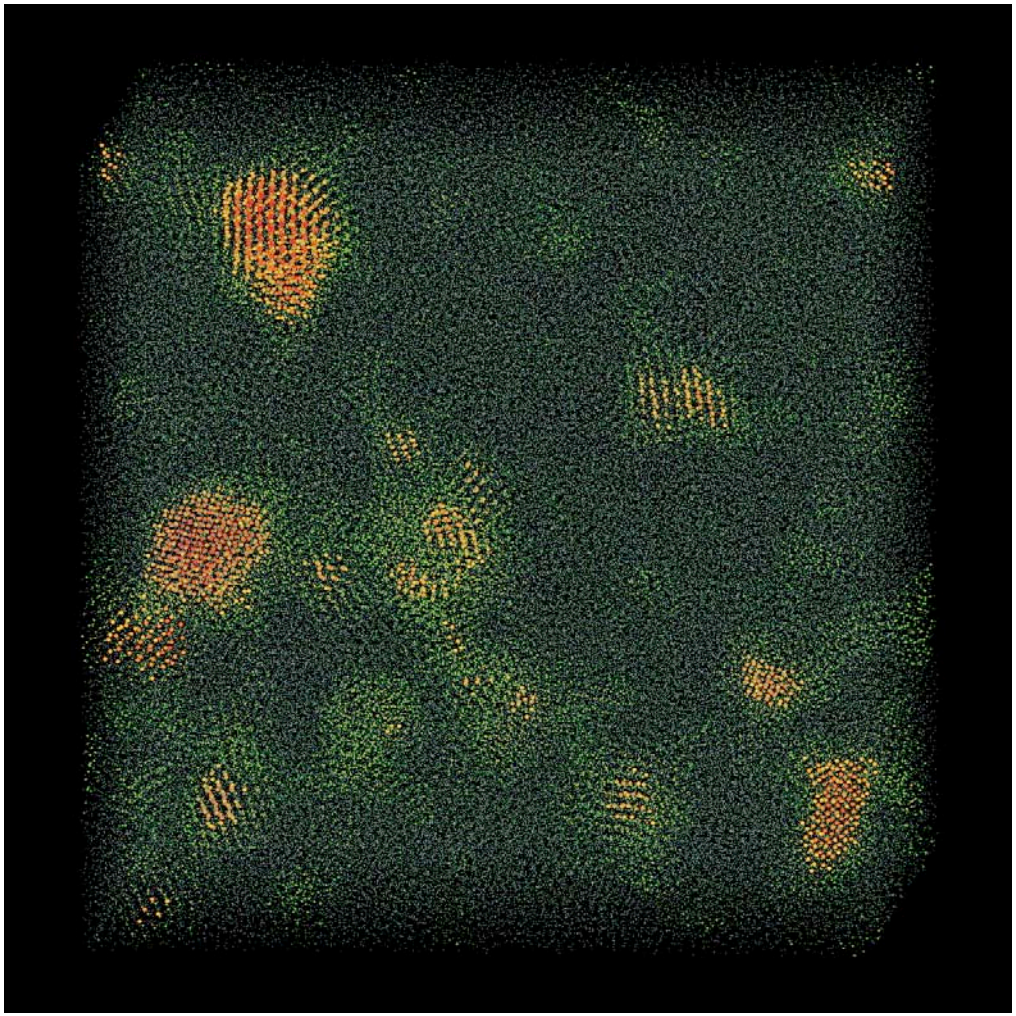
- Integration, integration, integration
- Test, test, test
- System will definitely run Pu aging codes in classified environment – Q3 2006?
- We want a 1 rack unclassified system but no funding has been identified
- Pursuing approval for swinging the machine from classified to unclassified
  - If approved we plan to swing the compute and IO nodes about 3-4 times per year

Ground Breaking Science

"You are looking at 256,000 tantalum atoms, with size and color determined by a local order parameter which has picked out several regions which are solidifying. The simulations starts with cell of molten Ta at 5000K and is then compressed (not frozen) into being a solid at about 2 Mbar of pressure. Most of the atoms are grey-green and very small, varying smoothly to red and large for highly ordered atoms. Displayed in this way, you can readily see the regions that are ordered. These nucleation sites have formed spontaneously out of the melt and will continue to grow, eventually forming a solid chunk of Ta with various grain boundaries. We've simulated the whole process.

"What Blue Gene has enabled us to do is perform this simulation on a large enough sample to see multiple such nucleation sites, so that we can (a) better understand how they form and (b) study how they interact as they grow. Previous simulations were limited to several thousand atoms, which corresponds to a cell size roughly the size of one of those nucleation sites. Needless to say, this has severely limited our ability to study the larger problem. It should be pointed out that many people have performed simulations using computationally cheap pair-wise potentials involving millions or even billions of atoms - but this is the first simulation on this scale using the very accurate but expensive MGPT potential, which is a quantum based potential."

Frederick H. Streitz
LLNL Rapid Resolidification Project



Summary

- We have done a lot of work and it's looking pretty good
- We aren't quite half way there
  - We have a long way to go
- This is the hardest thing we've ever attempted and we are right in the middle of it!